QA
76
M94
SI Ref.

M U D P I E no. 21

Museum and University Data, Program and Information Exchange

SURVEY OF COMPUTER PROGRAM CATALOGS-III

"COMPUTERS IN THE ENVIRONMENTAL SCIENCES"

This is an annual list of computer programs available for workers in the environmental sciences, with strong emphasis on geography. It is prepared by John R. Tarrant of the University of East Anglia, who states "the object...is to promote the sharing of information on the use of computers and the availability of computer programs in the environmental sciences." Publication is in January of each year, and entries for the issue must be received by September 30 of the previous year. The first issue, for 1971, includes material submitted by 249 individuals or departments, with the material arranged alphabetically by the names of the submitters. The programs available from each submitter are then numbered sequentially within his number, and may include as many as 52 listed programs (Dept. of Geography, University of Bristol, England). Each listing includes the following information: Name of submitter; Full Official Address; Computer used; Core size of machine and core required by program (if known); Peripherals needed; Language of program; Author of program; Address of author; Title and brief description of program; Purpose for which program is being used; and Keywords for indexing. Dr. Tarrant is interested in the receipt of programs for inclusion in future issues, and would like to have the above information for any program submitted. Users of the compilation are expected to write to the author for a copy of his program. An Index is included with fairly broad headings to guide one to material of interest.

The listing is published by and can be purchased for $2.00 from:

Geo Abstracts
School of Environmental Sciences
University of East Anglia
Norwich, NOR 88C, England

SMITHSONIAN
DEC 16 1971
LIBRARIES

A second listing available from the same source is "Computers in Geography (1970)," which sells for $1.25. I have not seen a copy of it, and cannot give any information concerning it.--James A. Peters.

SOMETHING NEW IN '72

Starting with the first issue in 1972 MUDPIE will include a section devoted to the storage and retrieval of catalogue data

concerning museum specimens, an area receiving considerable attention in museums around the world (see reports in this issue on the automation of data in systematic collections and on the Bolus Herbarium system). As an individual scientist and systematist watching the development of such systems, I remain very doubtful about their ultimate success and their value as opposed to their costs, but it seems rather dog in the mangerish for me not to recognize the overall interest in and movement towards the activity in museums. I hope that inclusion of material concerning this activity in MUDPIE will lead to the presentation of views both pro and con, and perhaps form the basic framework for a thorough discussion of the movement at the ICSEB meetings in Boulder in 1973.--JAP.

REMARKS WORTH REPEATING

In a recent paper included in the volume "Expanding Use of Computers in the 70's," edited by Gruenberger, Dick H. Brandon turns several felicitous phrases worth passing on to you, such as:

"Consider our overkill capacity in terms of printed output. Every computer has at least one high-speed printer, capable of producing 1000 pages per hour in six copies. Figuring 600 hours per month, and 65000 machines, it figures out to around 1200 pages per month for every person in the country. Have you read your 1200 pages this month? If we extrapolate to 1980 (300000 computers, printing speeds of 3000 lines per minute), the figure goes to 14,000 pages per person per month."

"Documentation is like sex: when it's good, it's very very good; and when it's bad, it's better than nothing."

"An industry of a million people, $22 billion in annual expenditures, and 65,000 machines, with a controversy over whether we slash the oh or slash the zero is totally ludicrous."

"I can go to work for you...as a systems analyst and do absolutely nothing for 18 months and you will not know it ...After six months, if you ask me, I'll tell you that I've been gathering data. Six months later I'll tell you I've been doing data analysis. And at the end of 18 months, I'll tell you that I've had to gather new data because the users have made changes. And if, after that, you challenge me to produce something, I'll quit."

"We have made unemployment totally obsolete. A programmer who is fired can get a n new job within two hours, and at an increase of $1000 a year (a number that can be called the loyalty gap). The loyalty gap for systems analysts is $2000 a year--and that is how you tell those two groups apart." (on this one, remember, he gave his talk in March, 1970!)

GENERAL COMMENTS ON THE DATA-BANKING SYSTEM FOR SPECIMEN INFORMATION DEVELOPED AT THE BOLUS HERBARIUM

The system referred to below was prepared with the aim of allowing the fullest possible use of the many kinds of information on specimen labels in taxonomic collections. Data are entered from specially formatted positions on punched cards. Simple verbal controls punched onto cards provide for data checking and input, selective retrievals, sequencing, listings, frequency analysis and map plotting.

1. APPLICATIONS: The system can be used for plant and animal collections from terrestrial or marine environments. It could readily be adapted for fossil or petrological data. It may serve as an aid in making taxonomic revisions, preparing local checklists, aiding in making identifications and setting up reference material for collecting expeditions. More uncommon uses may range from preparing lists of Types to tracing early collectors' itineraries.

2. DATA: The record for each specimen is formatted to include the following data: taxonomic position, species' name and author; type status, institution, collector and specimen number; ecology, frequency, altitude and date collected; the name of the locality and its geographical grid reference. In a modest-sized computer (16K core, 512K disk) there may be some 4,500 specimen records in a given bank. Individual banks may have several times more records than this, in a larger machine. Besides the main banks, files are maintained in the computer for authors', collectors' and institutions' names, citations for grid regions, ecology, frequency, Type status and the retrieval criteria.

3. DATA ENTRY: All entry operations are accompanied by printouts that can be checked and verified. The specimen records can be processed by an editing program. The printout from this can be used for proof-reading against the specimen labels. The cards are corrected before the data are admitted to storage in the bank. Facilities are provided for later changes to the stored records. To reduce punching effort, the taxonomic data for a species need only be entered on a single card. The taxonomic data are united with the specimen information on other cards (one per specimen), in the computer, to give the respective records. As in all other parts of the system, the user is aided by full sets of error messages for unacceptable data or instructions.

4. RETRIEVAL AND OUTPUT: The criteria for the selective retrieval of data from the bank may be broadly framed, or quite specific. A criterion, describing the subject (e.g. altitude) can be followed by a single datum, a list, or a range of data (e.g. 1000 m; 800 m, 900 m, 1000 m; 800--1000 m). Up to 40 criteria, each with a single datum, list or range, may be specified.

To qualify for retrieval, a record must match one of the data for every criterion cited. The retrieved records can then be set in an order given by further criteria. A useful ordering is by geographical co-ordinates and collectors. Classes of data in the record can be chosen to appear in the print-out of the retrieved information. In addition, individual kinds of data can be specified for listing in alphamerical sequence, with counts of the numbers of records in which each citation appears. Where geographical co-ordinates are specified for this, a distribution map of the collecting sites is given on a computer-linked plotter.

5. DATA COMPILATION AND RUN TIMES: Every effort has been made to strike a balance between the needs of both simplicity and shortening the time for compiling the raw data. These factors must also be seen in the need for avoiding over-long run times with modest computing facilities. Records with an average amount of data can be compiled at a rate of about 30 per hour, and proof-read at about 120 per hour. An IBM 1130 (16K core, 512K disks) took, for 1000 records, 39 minutes for full editing, 16 minutes for entry to the bank, 3 minutes for a simple retrieval operation and 4 1/2 minutes for plotting a small distribution map. These compilation and run times appear to be reasonably acceptable for a comparatively modest-sized machine.

A FORTRAN-IV program listing will be published shortly, and a general account of the system has been accepted for publication in an early issue of TAXON. At present the system is being adapted for a UNIVAC 1106 with remote-access and time-sharing facilities. Further inquiries from readers of MUDPIE would be welcome.--A. V. Hall, Asst. Curator, Bolus Herbarium, University of Cape Town, Rondebosch, C.P., Africa.

AUTOMATION OF DATA IN SYSTEMATIC COLLECTIONS

The Conference of Directors of Systematic Collections, which includes 31 representatives of most of the larger institutions in the United States and Canada, established at its meeting in May, 1970, a committee to study automation of data concerning specimens in natural history collections. This committee, composed of R. S. Cowan, Chairman (Smithsonian); Roland Force (Honolulu); N. G. Hairston (Michigan); and P. S. Humphrey (Kansas), reported to the 1971 meeting of the Conference. The group is primarily concerned with data "about" specimens, not in data derived from the study of the specimens. They are concerned not only with improvement of the ways of collecting and distributing this type of data, but also with ways to integrate inter-museum objectives in a cohesive plan. The Committee has concluded that both purposes are desirable, and has indicated certain stipulations and goals for immediate attention. They will attempt to develop format in such a way that restrictions on computers, systems (and hopefully, languages! --ed. note) will be minimal. They are interested in inter-museum

cooperation from the outset, to prevent incompatibility of files. They recommend establishment of a central Clearinghouse to assist participating museums, and they also recommend that the National Science Foundation be approached for funding.

The Committee recommended the Smithsonian as a temporary site for the Clearinghouse, with review and a decision as to a permanent location after three years of operation. They also recommended that a Board of Governors composed of the Directors of all participating Museums be established to determine Clearinghouse policy and act as liaison with the various concerned bodies. The full conference voted to accept and implement all of the recommendations of the committee.

Further developments concerning this activity will be communicated as they become available to MUDPIE.--JAP.

RECENT LITERATURE

Beaman, J. H., Ed. Some Applications for the Taxonomic Data Matrix: Six term papers by students at Michigan State University. Flora North America, report 63, 1971:1-103. (The six papers presented in this summary are listed below.)

Furlow, J. J. A computer program for analyzing native tree-species composition of local floras. FNA Rept. 63, 1971: 9-18.

Gage, S. H. A systematic approach to cereal leaf beetle behavior. FNA Rept. 63, 1971: 19-32.

Natarella, N. J. A computer data package for the selection of ornamental shrubs. FNA Rept. 63, 1971: 33-42.

Stevens, W. D. The general taxonomic data matrix format used for Sarcostemma (Asclepiadaceae). FNA Rept 63, 1971: 43-66.

Taggart, R. E. Automation of identification procedures in palynology using taxonomic data matrices. FNA Rept. 63, 1971: 67-90.

Valasek, D. M. A computerized phemological study of Sanford Natural Area. FNA Rept. 63, 1971: 91-103.

Berztiss, A. T. Data Structures--Theory and Practice. Academic Press, 1971: xiii + 442.

Dawson, C. B. & T. C. Wool. From Bits to If's--An Introduction to Computers and Fortran IV. Harper & Row, 1971: xii + 159. (Another publisher attempting to capture a little of the teaching market. Hardly worth the trouble of charging it from the library.--JAP)

Furlow, J. J., L. E. Morse, & J. H. Beaman. Computers in biological systematics, a new university course. Taxon, vol. 20, 1971: 283-290.

Gruenberger, F., Editor. Expanding Use of Computers in the 70's: Markets-Needs-Technology. Prentice-Hall, 1971: viii + 120.

Jackson, D. M. Classification, Relevance, and Information Retrieval. IN: Alt, F. L. & M. Rubinoff, Advances in Computers, vol. 11, 1971: 60-125.

Mitchell, R. K. Information Sciences and Computer Basics--an Introduction. Linnet Books, 1971: xii + 101. (A programmed text--does everyone find these as irritating to read as I? --JAP.)

Patten, B. C. Systems Analysis and Simulation in Ecology, Volume 1. Academic Press, 1971: xv + 607. (A series of contributed papers, divided into three parts: Introduction to Modeling; One-Species Models; and The Ecosystem: Simulation.)

Pylyshyn, Z. W. Perspectives on the Computer Revolution. Prentice-Hall Series in Automatic Computation, 1970: xx + 540. (A collection of readings concerning computers and their "relationships" to man. Everybody who is anybody is there, and they tend to repeat each other much more noticeably than if you were to read them separately and independently.--JAP.)

December, 1971
Division of Reptiles and Amphibians
National Museum of Natural History
Washington, DC 20560